Proc. Natl. Acad. Sci. USA
Vol. 78, No. 6, pp. 3408–3412, June 1981
Biochemistry

Zinc environment and *cis* peptide bonds in carboxypeptidase A at 1.75-Å resolution

(enzyme catalysis/proteases/protein crystallography)

D. C. REES*, M. LEWIS*, R. B. HONZATKO*, W. N. LIPSCOMB*†, AND K. D. HARDMAN‡

*Gibbs Chemical Laboratory, Harvard University, Cambridge, Massachusetts 02138; and ‡International Business Machines, Thomas J. Watson Research Center, Yorktown Heights, New York 10598

Contributed by William N. Lipscomb, February 27, 1981

ABSTRACT **The structure of the metalloenzyme carboxypeptidase A (peptidyl-L-amino-acid hydrolase, EC 3.4.17.1) has been refined at 1.75 Å by a restrained least-squares procedure to a conventional crystallographic *R* factor of 0.162. Significant results of the refined structure relative to the catalytic mechanism are described. In the native enzyme, the zinc coordination number is five (two imidazole Nδ1 nitrogens, the two carboxylate oxygens of glutamate-72, and a water molecule). In the complex (at 2.0-Å resolution) of carboxypeptidase A with the dipeptide glycyl-L-tyrosine, however, the water ligand is replaced by both the carbonyl oxygen and the amino nitrogen of the dipeptide. The amino nitrogen also statistically occupies a second position near glutamate-270. Consequently, the coordination number of zinc may vary from five to six in carboxypeptidase A–substrate complexes. Implications of these results for the catalytic mechanism of carboxypeptidase A are discussed. In addition, three *cis* peptide bonds, none of which involves proline as the amino nitrogen donor, have been located fairly near the active site.**

Proposed mechanisms (1) for the hydrolysis of peptide and ester substrates by carboxypeptidase A (CPase A; peptidyl-L-amino-acid hydrolase, EC 3.4.17.1) have relied heavily on the crystal structure of this enzyme for detailed molecular interpretations. CPase A_α is a zinc-containing metalloenzyme which has 307 amino acid residues. During the first stages of substrate hydrolysis, the zinc presumably polarizes the carbonyl group at the scissile bond; glutamate-270 acts as either a nucleophile or a general base catalyst. Tyrosine-248 (or a water molecule) may function as the proton donor. No single catalytic pathway has accounted for the hydrolysis of all substrates. Recent studies of ester hydrolysis have focused on nucleophilic attack to yield the anhydride intermediate and on possible changes in coordination of the zinc.

In order to evaluate alternative mechanisms, precise characterization of the active site of CPase A is essential. Consequently, we have undertaken refinement of the CPase A structure in the native state and in complexes with dipeptides (2–4), with substrate analogues (5), and with the 39-amino acid inhibitor of CPase A from the potato (6).

In this paper, we describe several of the most striking results from the refinement of the native CPase A structure at 1.75-Å resolution. [The solution of the native CPase A structure to 2.0-Å resolution has been documented (2–4).] Of particular relevance to the catalytic mechanism is the change in zinc coordination number from five in the native state to five or six in the complex with glycyl-L-tyrosine (Gly-Tyr). In addition, three *cis* peptide bonds, none of which involves proline as the amino nitrogen donor, have been located near the active site. We comment on the relationships of these results to catalysis of substrate hydrolysis and metal ion exchange.

MATERIALS AND METHODS

Bovine CPase A (Cox) was purchased from Sigma and used without further purification. Crystals for the collection of x-ray diffraction data in the range 2.0–1.75 Å were prepared by dialysis of CPase A against 0.2 M LiCl/20 mM Tris·HCl, pH 7.5. The unit cell dimensions of these crystals (a = 51.60 Å; b = 60.27 Å; c = 47.25 Å; β = 97.27°; space group $P2_1$) agree to within 0.5% with the lattice constants of the crystals used in the earlier work. This crystal form is distinct from a second crystal form which is characteristic of CPase A (Anson) (7). Although some difficulties in obtaining the "x-ray crystals" from commercial CPase A have been reported (8), we readily obtained excellent single crystals.

The 2.0- to 1.75-Å data set was collected from a single crystal on a Syntex $P2_1$ diffractometer at 4°C. Decay in intensities of four check reflections was less than 20% at the end of data collection. The data were corrected for absorption (9) and Lorentz-polarization effects. Scaling to the original 2.0-Å data set was accomplished by collecting the entire h0l zone to 1.75 Å. Merging of the two data sets gave an R factor, $\Sigma\,(|I_1 - I_2|)/\Sigma\,(I_1 + I_2)$ in which I_1 and I_2 represent equivalent reflections observed in the two data sets, of 0.09. The final data set from 8.0- to 1.75-Å resolution consisted of 21,530 reflections whose structure factor amplitude exceeded twice the estimated standard deviation.

The coordinates for the CPase A model obtained from the 2.0-Å multiple isomorphous replacement map had originally been idealized and refined by using the model building and real space refinement programs of Diamond (10, 11). Of particular interest to the present work was the earlier description of the zinc coordination as tetrahedral (two Nδ1 nitrogens from histidine-69 and -196, a carboxylate oxygen from glutamate-72, and a water molecule) and the presence of a *cis* peptide bond between serine-197 and tyrosine-198. These coordinates provided the starting model for refinement in the present work using the Hendrickson–Konnert restrained least-squares algorithm (12, 13). In this program, coordinates and temperature factors for the individual atoms are varied to minimize the residual $\Sigma\,(|F_o| - |F_c|)^2$, in which F_o and F_c are the observed and calculated structure factors, respectively, and the summation is over all observed reflections. To increase the number of observations per refined parameter, the least-squares observational equations are supplemented with additional terms which aid in maintaining proper stereochemistry in the model.

Details of the refinement will be described elsewhere. In brief, the starting CPase A coordinate set was subjected to 31

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviation: CPase A, carboxypeptidase A.

† To whom reprint requests should be addressed.

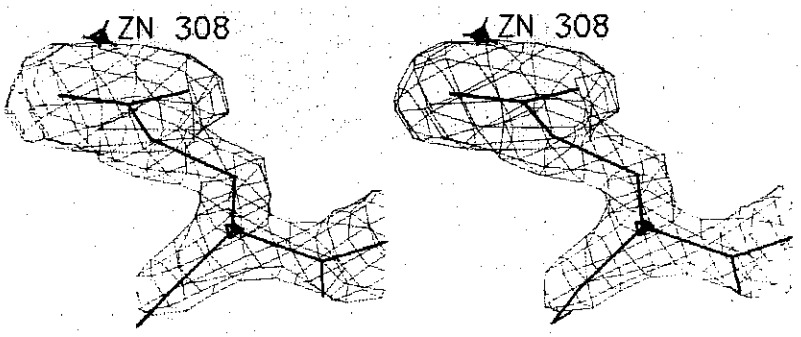

FIG. 1. Region around glutamate-72 from an ($|F_o| - |F_c|$) difference map using calculated phases, in which glutamate-72 was omitted from the structure factor calculation. The final refined position of glutamate-72 and the zinc are superimposed on the difference map. The placement of both carboxylate oxygens of glutamate-72 on the zinc is shown. We count both oxygens of glutamate-72 in this study; an alternate view is that this is a variant of the binding mode of a single ligand.

cycles of coordinate and isotropic temperature factor refinement, resulting in a final model with 2664 atoms (2437 atoms other than H in the 307 residues of CPase A, 1 zinc, and 226 water molecules). After every five to eight refinement cycles, difference maps were calculated by using Fourier coefficients ($|F_o| - |F_c|$) and ($2|F_o| - |F_c|$) and refined phases from the model in order to check for misbuilt residues and for solvent molecules. Rebuilding of the CPase A model was accomplished with the use of Diamond's program BILDER, as extensively modified and adapted by R. C. Ladner for a VAX 11/780 and an Evans and Sutherland Picture System II. During the course of the refinement, the crystallographic R factor ($\Sigma \; ||F_o| - |F_c||)/\Sigma \; F_o$ decreased from 0.366 for the 14,179 reflections between 8.0- and 2.0-Å resolution to 0.162 for the full data set at 1.75-Å resolution. The stereochemistry of the final CPase A model was quite reasonable: the average root mean square deviation from ideal bond distances and angles were 0.024 Å and 3.5°, respectively. The average shift from the starting coordinate set was 0.73 Å (0.49 Å for main chain atoms; 0.91 Å for side chain atoms).

RESULTS AND DISCUSSION

The smoothness of the refinement procedure and the absence of major rebuilding of residues during refinement testify to the accuracy of the original CPase A model. Nevertheless, increased resolution and improved phases have permitted clarification or revision of several aspects of the CPase A structure. We describe below several new features near the active site of CPase A which may be significant for understanding the catalytic mechanism.

The zinc coordination in CPase A was reported originally as a distorted tetrahedron. Difference maps calculated during the initial stages of the present refinement suggested that the glutamate-72 carboxylate group had been slightly misplaced and that both oxygen atoms of the carboxylate anion are coordinated to the zinc. Subsequent refinement cycles verified this change. These refinements yielded much lower temperature factors for the carboxylate group in its revised position. An ($|F_o| - |F_c|$) difference Fourier map calculated with structure factors from a model in which glutamate-72 was omitted confirmed the placement of both carboxylate oxygens on the zinc (Fig. 1). Consequently, the coordination number of the zinc in native CPase A is five. Zinc–ligand bond distances and representative values for zinc–amino acid complexes are given in Table 1.

On the basis of spectroscopic and magnetic susceptibility studies of cobalt(II)-substituted CPase A, which is similar to Zn-CPase A, Gray and coworkers postulated borderline four to five coordination for the metal center (15, 16). The present results are entirely consistent with this proposal. NMR studies of ^{19}F in solutions of Mn-CPase A indicated that both a fluoride ion and at least one water molecule were coordinated to manganese (17); if glutamate-72 also binds to manganese(II) with both oxygens, the metal ion coordination number would be six. However, the coordination geometry of manganese is not always like that of zinc. Thus, the applicability of these results to Zn-CPase A is open to question.

In order to examine zinc coordination in a peptide–CPase A complex, a new map of the complex of CPase A with Gly-Tyr was calculated at 2.0-Å resolution. The Gly-Tyr–CPase A map was generated by using coefficients ($6\,|F_{gt}| - 5|F_{nat}|$) and calculated phases, in which F_{gt} and F_{nat} are structure factors for the Gly-Tyr complex and native enzyme, respectively. Because the Gly-Tyr complex is approximately 30% occupied, this map is roughly equivalent to a more traditional ($2|F_{gt}| - |F_{nat}|$) map, which would be appropriate if the complex were fully occupied. As shown in Fig. 2, the carbonyl oxygen and amino nitrogen of Gly-Tyr are bound to the zinc in addition to the four atoms of the protein ligands. Consequently, the coordination number for the zinc can be as high as six in the Gly-Tyr complex.

Interestingly, the electron density map gives evidence that the amino nitrogen statistically occupies a second position near the carboxylate group of glutamate-270. The two conformations of the amino nitrogen are related by a rotation of approximately 180° about the glycine α carbon-carbonyl carbon bond. Based on values of the electron density at both positions, the amino nitrogen appears to be approximately equally distributed between both sites. If the α carbon occupies the same position in both conformations, it appears unlikely that a water molecule could bind to the zinc when the amino nitrogen is oriented toward glutamate-270. The coordination number of zinc is then five in the complex for this alternate orientation. There is no evidence in the electron density map for alternate positions for the remainder of the Gly-Tyr molecule. The electron density for glutamate-270 is more diffuse than expected, however, possibly indicative of alternate conformations.

Changes in coordination number of zinc have also been observed in the binding of inhibitors to human carbonic anhydrase (18) and horse liver alcohol dehydrogenase (19). In these two cases, the coordination state changes from four to five. Although common in complexes of small molecules with zinc, six-coordinated zinc as in this CPase A complex has not been reported in metalloenzymes.

The demonstration of six-coordinated zinc has important im-

Table 1. Zinc–ligand bond distances in native CPase A structure

Ligand	Distance, Å*	
	r_{obs}	r_{mod}
His-69 Nδ1	2.10	2.00–2.20
His-196 Nδ1	2.08	2.00–2.20
Glu-72 Oε1	2.23	2.15
Glu-72 Oε2	2.33	2.58
H_2O	1.96	2.11

* r_{obs}, Observed bond distance; r_{mod}, representative bond distances for zinc–amino acid complexes, taken from ref. 14. Estimated errors in the bond lengths are approximately 0.2 Å.

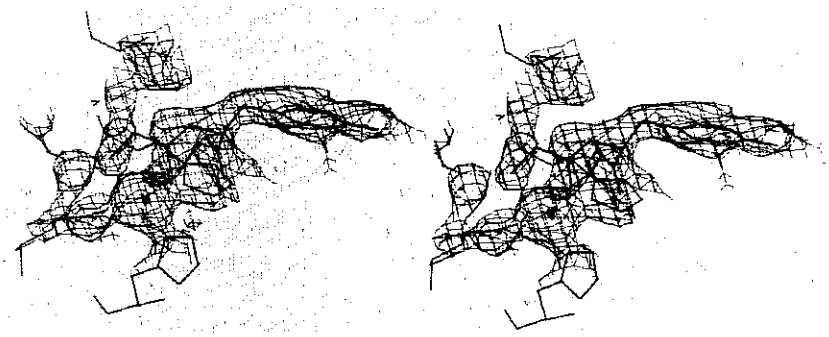

FIG. 2. Coordinates for Gly-Tyr and selected active site residues of CPase A are superimposed on a difference map of the Gly-Tyr–CPase A complex described in the text. The zinc position is marked with a tetrahedron: glutamate-270 is at the top of the figure. The zinc ligands from left are glutamate-72, histidine-69, and histidine-196. Note the density for the amino nitrogen of Gly-Tyr extending toward both the zinc and glutamate-270.

plications for the catalytic mechanism of CPase A. A hydroxide ion occupying a sixth coordination site would be ideally situated to deacylate anhydride intermediates which are also coordinated to the zinc. Makinen *et al.* (20) have recently obtained electron paramagnetic resonance evidence that such a situation exists in cobalt(II)-CPase A for the ester *o*-(*p*-chlorocinnamoyl)-L-β-phenyl lactate. In addition, attempts to detect the presence of anhydride intermediates through isotope exchange reactions may be misleading if the water molecule produced during anhydride formation binds directly to the zinc and is not available to exchange freely with the solvent. This ambiguity was recognized by Breslow and Wernick (21).

The observed structure of the complex, together with general aspects of the catalytic mechanism of CPase A, provide plausible explanations for the observation that Gly-Tyr is a poor substrate. When the amino nitrogen binds to the zinc and displaces the zinc-bound water (the nitrogen is at the site of the bound water in the native enzyme), then the water might be unable to participate in hydrolysis until it displaces the amino nitrogen. Acylated dipeptides, or longer substrates, would not show this anomaly. Conversely, when the amino nitrogen of Gly-Tyr binds to glutamate-270, this residue would be hindered from acting as either a nucleophile or a general base catalyst, thus decreasing the rate of hydrolysis.

Several lines of evidence independent of the crystallographic studies suggest that the amino nitrogen of Gly-Tyr may be coordinated to the zinc. On the basis of the pH dependence of the inhibition kinetics of CPase A by Gly-Tyr, Yanari and Mitz (22) proposed that the anionic form of Gly-Tyr binds to the enzyme. Electrostatic interactions would favor unprotonated nitrogen binding to the zinc and protonated nitrogen binding to the carboxyl group of glutamate-270. As the pH is increased from the value of 7.5 used in the present crystallographic work, an increasing fraction of the amino nitrogen should be bound to the zinc. In addition, Makinen *et al.* (20) observed that binding of Gly-Tyr to cobalt(II)-CPase A displaced the metal-bound water in their electron paramagnetic resonance study. Although it is possible that binding of the carbonyl oxygen alone could displace the water, the proposed position of the amino nitrogen corresponds closely to the site of the bound water in the native enzyme (Fig. 3).

The binding of Gly-Tyr to CPase A is accompanied by displacement of several water molecules, as is clearly indicated in Fig. 3. This map showing water structure was generated by using difference Fourier coefficients ($|F_o| - |F_c|$) and calculated phases in which all water molecules were omitted from the structure factor calculation. The conformations of selected active site residues in the native enzyme are also included for reference. An important question is whether the enzyme can make use of these changes of hydration in catalysis. Low and Somero (23) have explained enzyme activation or inhibition by neutral salts by proposing that parts of the protein change their exposure to water during conformational changes in catalysis. Conformational changes during substrate binding, hydrolysis, and product dissociation stages can easily be influenced by changes in solvent exposure of enzyme and substrate. The enzyme could well select for backbone (deletions, additions) and side chains which would make use of the solvent interactions to facilitate the reaction. These effects may apply particularly to the binding and catalytic groups of the enzyme and substrate and perhaps even to much larger regions of the enzyme. However, specific structural models for these phenomena are not currently available. Studies of hydration in electron density maps at high resolution may contribute to an understanding of these effects.

One interesting aspect of protein structures is the relative absence of *cis* peptide bonds observed in most highly refined protein structures. The energetic difference between *cis* and *trans* peptide bonds not involving proline as the nitrogen donor leads to the expectation that approximately 1% of these peptide bonds should be *cis* (24). However, the incorporation of most peptide bonds into primary, secondary, and tertiary structure can be expected to give a much smaller fraction of bonds that

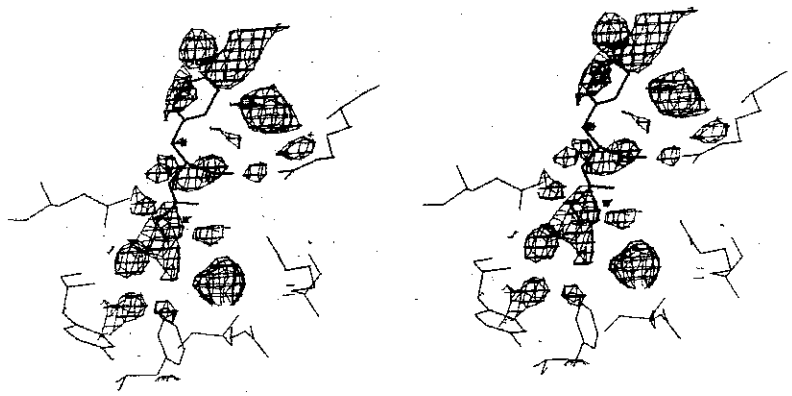

FIG. 3. Water structure in the active site of CPase A as seen in a difference map calculated as described in the text. The Gly-Tyr coordinates superimposed on the map indicate the extent of desolvation which must occur when Gly-Tyr binds to CPase A. Native conformations of the following active site residues of CPase A are indicated, clockwise from the upper right: arginine-145, arginine-127, arginine-71, phenylalanine-279, tyrosine-198, and glutamate-270. The zinc position is marked by the point at the center of the figure.

isomerize to the *cis* configuration (25). Although only one *cis* peptide bond, between serine-197 and tyrosine-198, had been described in the original CPase A model, the presence of two additional *cis* bonds (between proline-205 and tyrosine-206, and between arginine-272 and aspartate-273) was unexpectedly observed during refinement of CPase A with the use of the Agarwal fast Fourier least-squares algorithm (26). These two bonds were initially restrained in the *trans* conformation at the start of our use of the Hendrickson–Konnert refinement. Nevertheless, the ω angle (twist) of these bonds shifted drastically from 180° (*trans*) during the refinement. Consequently, these bonds were restrained to the *cis* configuration in all subsequent refinements. In Fig. 4 we show the region around each *cis* peptide bond in a difference map calculated upon omission of these six residues from the structure factor calculation. The *cis* nature of each bond is clearly indicated. Although only one of the six residues (tyrosine-198) has been implicated in the extended active site region, all of these bonds are located fairly near the active site. Interestingly, they also are located immediately adjacent to the carboxy terminus of β-strands. Moreover, values for the conformational angle ψ of the first residue are approximately $-110°$, whereas values for the conformational angle ϕ of the second residue are approximately 160°.

No functional properties have been associated with these *cis* bonds. From a NMR investigation of the kinetics of metal binding to concanavalin A, it was suggested that a *cis–trans* isomerization accompanies metal binding (27). Two independent crystallographic analyses of the structure of demetallized concanavalin A have produced conflicting results concerning the existence of the postulated isomerization: in one study at 2.8-Å resolution the isomerization was observed (28); these results were questioned in a later study at 3.2-Å resolution (29). Structure refinement at higher resolution should resolve this discrepancy. So far, no correlation of *cis–trans* peptide bond isomerization with metal binding has been observed with CPase A. However, Zisapel (30) has found a correlation of kinetics of nitration of zinc-reconstituted apo-carboxypeptidase B with the length of time that the enzyme stayed in the zinc-free state. It has been suggested that a *cis–trans* peptide bond isomerization may account for these results (1).

With the exception of the binding of the amino terminus of Gly-Tyr statistically either to the zinc or to the carboxylate group of glutamate-270 in its complex with CPase A, previous conclusions concerning the nature of the catalytic groups and the proposed substrate binding modes have not been altered by the findings in the present study. Demonstration of an additional zinc ligand (aside from those to the enzyme) is consistent with the idea that a water molecule or hydroxyl ion may bind to zinc during a later stage of catalysis. The role of a zinc-bound water or hydroxide ion in deacylation of anhydride intermediates was suggested earlier (31), although only for four-coordinated zinc, and has been demonstrated spectroscopically for

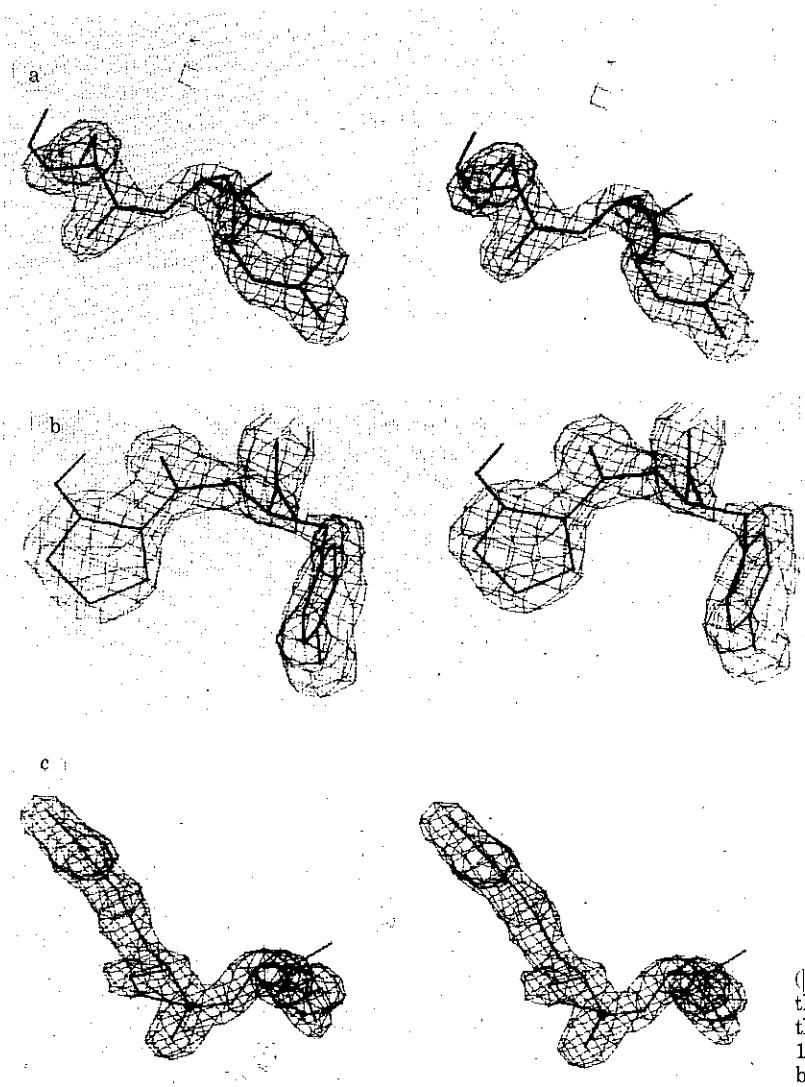

FIG. 4. Fit of the *cis* peptide bonds of CPase A to an $(|F_o| - |F_c|)$ difference map with calculated phases, in which the six residues involved in these bonds were omitted from the structure factor calculation. (*a*) Serine-197 and tyrosine-198 peptide bond; (*b*) proline-205 and tyrosine-206 peptide bond; (*c*) arginine-272 and aspartate-273 peptide bond.

certain ester substrates with which the coordination of zinc increases by one in the reaction (20).

We thank R. C. Ladner for his assistance with the graphics programs, the National Institutes of Health (GM 06920) for support of this research, the National Science Foundation (PCM 77-11398) for support of the computational facilities, and the American Cancer Society for a Postdoctoral Fellowship (M.L.).

1. Lipscomb, W. N. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3875–3878.
2. Lipscomb, W. N., Hartsuck, J. A., Reeke, G. N., Jr., Quiocho, F. A., Bethge, P. A., Ludwig, M. L., Steitz, T. A., Muirhead, H. & Coppola, J. C. (1968) *Brookhaven Symp. Biol.* **21**, 24–90.
3. Quiocho, F. A. & Lipscomb, W. N. (1971) *Adv. Prot. Chem.* **25**, 1–78.
4. Hartsuck, J. A. & Lipscomb, W. N. (1971) in *The Enzymes*, ed. Boyer, P. (Academic, New York), Vol. 3, pp. 1–56.
5. Rees, D. C., Honzatko, R. B. & Lipscomb, W. N. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3288–3291.
6. Rees, D. C. & Lipscomb, W. N. (1981) *Proc. Natl. Acad. Sci. USA* **77**, 4633–4637.
7. Quiocho, F. A. & Richards, F. M. (1964) *Proc. Natl. Acad. Sci. USA* **52**, 833–839.
8. Johansen, J. T. & Vallee, B. L. (1975) *Biochemistry* **14**, 649–660.
9. North, A. T. C., Phillips, D. C. & Mathews, F. S. (1968) *Acta Crystallogr. Sect A.* **24**, 351–359.
10. Diamond, R. (1966) *Acta Crystallogr.* **21**, 253–266.
11. Diamond, R. (1971) *Acta Crystallogr. Sect A* **27**, 436–452.
12. Konnert, J. H. (1976) *Acta Crystallogr. Sect A* **32**, 614–617.
13. Hendrickson, W. A. & Konnert, J. H. (1980) in *Biomolecular Structure, Function, Conformation and Evolution*, ed. Srinivasan, R. (Pergamon, Oxford), Vol. 1, pp. 43–57.
14. Freeman, H. C. (1967) *Adv. Prot. Chem.* **22**, 257–424.
15. Rosenberg, R. C., Root, C. A., Wang, R.-H., Cerdonio, M. & Gray, H. B. (1973) *Proc. Natl. Acad. Sci. USA* **70**, 161–163.
16. Rosenberg, R. C., Root, C. A. & Gray, H. B. (1975) *J. Am. Chem. Soc.* **97**, 21–26.
17. Navon, G., Shulman, R. G., Wyluda, B. J. & Yamane, T. (1970) *J. Mol. Biol.* **51**, 15–30.
18. Kannan, K. K., Petef, M., Fridborg, K., Cid-Dresdner, H. & Lovgren, S. (1977) *FEBS Lett.* **73**, 115–119.
19. Boiwe, T. & Branden, C.-I. (1977) *Eur. J. Biochem.* **77**, 173–179.
20. Kuo, L. C. & Makinen, M. W. (1979) *A Conference on Methods for Determining Metal Ion Environments in Proteins*, Jan. 10–12, 1979, Las Cruces, N.M., abstr. 27.
21. Breslow, R. & Wernick, D. L. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 1303–1307.
22. Yanari, S. & Mitz, M. A. (1956) *J. Am. Chem. Soc.* **79**, 1154–1158.
23. Low, P. S. & Somero, G. N. (1975) *Proc. Natl. Acad. Sci. USA* **77**, 3305–3309.
24. Ramachandran, G. N. & Sasisekharan, V. (1968) *Adv. Prot. Chem.* **23**, 283–437.
25. Ramachandran, G. N. & Mitra, K. (1976) *J. Mol. Biol.* **107**, 85–92.
26. Agarwal, R. C. (1978) *Acta Crystallogr. Sect A* **34**, 791–809.
27. Brown, R. D., Brewer, C. F. & Koenig, S. H. (1977) *Biochemistry* **16**, 3383–3396.
28. Reeke, G. N., Jr., Becker, J. W. & Edelman, G. M. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 2286–2290.
29. Shoham, M., Yonath, A., Sussman, J. L., Moult, J., Traub, W. & Kalb (Gilboa), A. J. (1979) *J. Mol. Biol.* **131**, 137–155.
30. Zisapel, N. (1978) *Biochem. Biophys. Res. Commun.* **81**, 28–34.
31. Lipscomb, W. N. (1974) *Tetrahedron* **30**, 1725–1732.